Proc. of the 2017 IEEE International Conference on Signal and Image Processing Applications (IEEE ICSIPA 2017), Malaysia, September 12-14, 2017

Transfer Learning For Diabetic Macular Edema (DME) Detection On Optical Coherence Tomography (OCT) Images

Genevieve C. Y. Chan, Awais Muhammad, Syed A. A. Shah, Tong B. Tang, Cheng-Kai Lu and Fabrice Meriaudeau
Centre for Intelligent Signal and Imaging Research Department of Electrical and Electronic Engineering,
Universiti Teknologi PETRONAS
32610 Bandar Seri Iskandar, Perak Darul Ridzuan, Malaysia.
genevieve.chan94@gmail.com, m.awais0100@gmail.com, engrayazshah@yahoo.com, tongboon.tang@utp.edu.my, chengkai.lu@utp.edu.my, fabrice.meriaudeau@utp.edu.my

***Abstract*— Diabetic Macular Edema (DME) is a common eye disease that causes irreversible vision loss for diabetic patients, if left untreated. Thus, early diagnosis of DME could help in early treatment and prevent blindness. This paper aims to create a framework based on deep learning for DME recognition on Spectral Domain Optical Coherence Tomography (SD-OCT) images through transfer learning. First, images are pre-processed: denoised using Block-Matching and 3-Dimension (BM3D) filtering and cropped through image boundary extraction. Later, features are extracted using CNN of AlexNet and finally images are classified using SVM classifier. The results are evaluated using 8-fold cross-validation. The experiments show that denoised and cropped images lead to better classification performances, exceeding previous other recent published works of 96% accuracy.**

Keywords: Diabetic Macular Edema (DME), Convolutional Neural Network (CNN), Spectral Domain Optical Coherence Tomography (SD-OCT).

I. INTRODUCTION

Diabetic Macular Edema (DME) is one the many eye diseases that is commonly found in diabetic patients, typically Type 2 diabetes. A review in 2012 shows that almost 7% of diabetic patients may have DME [1]. DME is a result of the accumulation of fluid in the macula due to Diabetic Retinopathy (DR) i.e. the damage of blood vessels in the retina of the eye. Consequently, an increase of retinal thickness within one disk diameter of the fovea center with or without hard exudates and sometimes associated with cysts [2]. Evidently, DME will only happen when DR is left untreated. When DME is unattended, it may cause blurry vision and if the worst comes to worst, an irreversible vision loss. Health care and associated costs related to eye disease is also very high when case becomes severe.

There is an early detection process for such diseases e.g. fundus images and Optical Coherence Tomography (OCT) to avoid the severe effects of DME. These methods basically capture images of the retina. An irregular thickness of the retina indicates the presence of DME. Increasingly, Spectral Domain OCT (SD-OCT) images are preferable over fundus images as it can capture images with high resolution of the thickness of the retina and thus of better quality for analysis [3]. An automated detection of DME from SD-OCT images brings us one step towards artificial intelligence technology. This study aims to formulate a framework based on deep learning for DME identification from SD-OCT images through transfer learning technique, given limited data resources, to provide a solution for better reliance in DME detection process.

The main aim of the paper is to investigate deep learning for DME recognition on OCT images. In Section II we review the literature on the use of deep learning. Section III discusses the methodology of the proposed algorithm. In section IV the results and discussions are presented and the paper is concluded in section V.

II. LITERATURE REVIEW

Over the last twenty years, lots of work has been done in the field of Artificial Intelligence (AI) i.e. a machine capable of behaving like a human e.g. the ability to feel and make decisions, was first introduced in 1950 [4]. In other words, researchers desire to create something that does something intelligently. The success of AI remains unanswered until 1980, when machine learning begins to flourish [4], and increasingly, deep learning after a historic breakthrough in the ImageNet Computer Vision competition in 2012, where deep convolutional networks trained on a dataset of a million images of one thousand classes produced significant decrease of error rates i.e. half of the previous competing approach [5].

A. Deep Learning

Machine-learning technology has progressively contributed to the modernization of the society e.g. object recognition, speech recognition and selection of relevant results while searching the web. It is widely available and has already been used in our everyday lives e.g. smartphones, cameras and web searches. The basic idea behind these technologies is called representation learning, which automatically identifies the representations of a given raw input by extracting features needed, using a feature extractor, to train a classifier to detect or classify patterns in the input. Consequently, these applications' systems evolve into deep learning e.g. the social network, Facebook's photo tagging feature is a face recognition based on a deep network called DeepFace [6].

Proc. of the 2017 IEEE International Conference on Signal and Image Processing Applications (IEEE ICSIPA 2017), Malaysia, September 12-14, 2017

B. *AlexNet Architecture*

Fig. 1 shows the architecture of AlexNet developed by Krizhevsky et al. with 5 convolutional layers and 3 fully connected layers [7]. In our approach, the probability values after the FCNs are used as features to feed a SVM classifier.

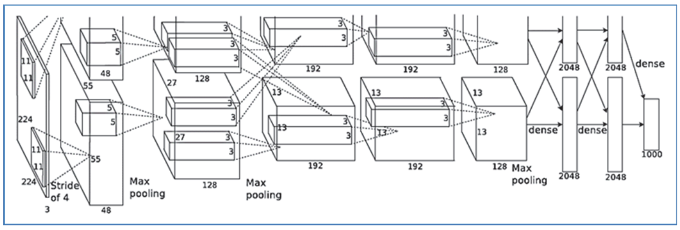

Fig. 1 An illustration of the AlexNet Architecture using CNN along with the filter banks value at each layer. Adapted from "ImageNet Classification with Deep Convolutional Neural Networks" by Krizhevsky A. et al. (2012) [7].

A recent review conducted by Massich et al [8] compares state of the art research methods on classifying DME and Normal patients using SD-OCT volumes. There are six researches reviewed and it is found that each research follows a similar process flow i.e. image pre-processing, feature detection, mapping feature representation and classification, but with different implementation in each stage. Note that in the present research, input images do not go through feature detection and mapping, feature representation are done by CNN and classification is done using a Support Vector Machine (SVM) classifier. Therefore, the comparative study will only focus on these areas.

III. Methodology

32 volumes of OCT images ($\boldsymbol{A} \in \mathbb{R}^{128x1024x512}$) and corresponding labels ($l \in$ {normal, dme}) are used as the input for the network. AlexNet is designed for 227×227 images with three channels (RGB images), thus all OCT images are first extracted from their volumes ($\boldsymbol{A} \in \mathbb{R}^{1024x512}$), pre-processed ($A \in \mathbb{R}^{227x227}$) and concatenated thrice ($\boldsymbol{A} \in \mathbb{R}^{3x227x227}$). Through transfer learning, features are extracted with CNN using the pre-trained weights trained on ImageNet dataset [7]. Finally, the features extracted are used to train a SVM classifier, with 8-fold cross-validation for all 4096 images. Accuracy (ACC), Specificity (SP) and Sensitivity (SE) are evaluated at the end of each fully connected layer to evaluate the performance of the results.

A. *Dataset*

The dataset of retinal images is obtained from the Singapore Eye Research Institute (SERI), using CIRRUS TM (carl Zeiss Meditec, Inc., Dublin, CA) SD-OCT device [9]. It contains 32 OCT volumes i.e. 16 volumes with DME diagnosed and 16 volumes normal cases. Each volume is captured with 128 B-scans with resolution of 1024px × 512px. All SD-OCT volumes are read and assessed by trained graders and identified as normal or DME cases based on evaluation of retinal thickening, hard exudates, intra-retinal cystoid space formation and sub retinal fluid [9]. Within the DME sub-set, many lesions have been selected to create a rather complete and diverse DME dataset.

B. *Method*

1) *Noise removal using BM3D filtering*

The retinal OCT images have fuzzy and unclear edges which are considered as noise. Thus, the images are smoothen using BM3D for pre-processing. BM3D is a denoising method in which a reference patch is group together with similar 2D patches into 3D groups. For more details about the implementation, please refer to Lebrun's work [10]. A comparative study shown that BM3D filtering has higher overall effect than other filtering method [11]. The images are read from each volume of OCT images. Then the images are resized, as the adopted AlexNet is designed for 227×227 input image size. Then, the images are concatenated thrice after BM3D filtering [12]. Fig. 2 shows the output of the BM3D filtering after noise reduction.

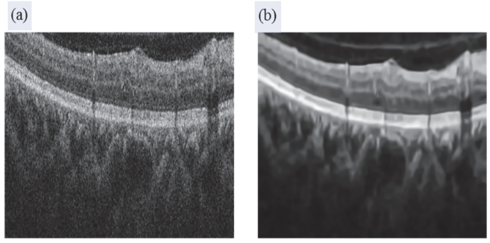

Fig. 2 BM3D Filtering. (a) Noisy image. (b) BM3D Filtered

2) *Image boundary extraction*

Fig. 3 shows the flowchart of image boundary extraction. The aim of this process is to detect the first layer and the final layer of the retina, which is then used to perform cropping. The process is adopted from Stephanie *et al.* [13], whereby each pixel of the image a.k.a. node are connected through the edge and a certain weight is assigned to each edge to create path preferences. The nodes with characteristics of the layers are assigned with low weights between the connected nodes. Then, using minimum total weight sum as the path preference, the retinal layer can be segmented and highlighted. Fig. 4 shows the output of the image boundary extraction.

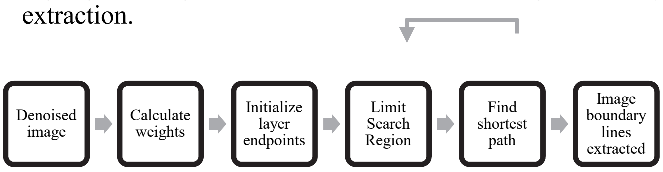

Fig. 3 A generalized image boundary extraction schematic [13]

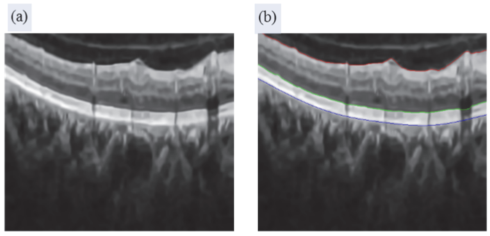

Fig. 4 Image Boundary Extraction. (a) BM3D Image without boundary lines. (b) BM3D Image with boundary lines: ILM – Red Line; ISOS – Green Line; RPE – Blue Line.

3) *Image cropping*

Note that the features that distinguish the differences between DME and Normal patients are the swelling of the retinal layer at the fovea. Consequently, the pixels which are not specific for normal or DME detection are therefore not relevant and should be removed. Using the output images of image boundary extraction stage, the image is cropped and stored. Fig. 5 shows the output of the result of image cropping.

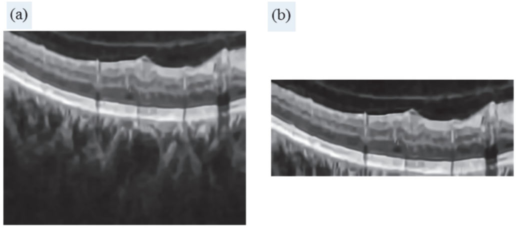

Fig.5 Image Cropping. (a) BM3D Image without cropping. (b) BM3D image after cropping

4) *Feature extraction*

Before the features extraction the dataset is divided into training set and test set, where 4096 images are partitioned into 8 subsamples of 512 equal sizes of images and 7 subsamples are used as training set and the remaining subsample is the test set. Then, the training features are extracted using CNN at different depths in the network after the Fully Connected Layers (FCLs).

5) *Image Classification using SVM classifier*

After all features are extracted, they are used to train a classifier to perform image classification. The classifier used is a SVM classifier [17], in which a hyperplane is drawn to distinguish between two classes (i.e. dme or normal). The classification is conducted on dataset with 4 different scenarios:

- Experiment #1 is carried out on raw datasets with no noise removal and no image cropping
- Experiment #2 is carried out on datasets with noise removal but without image cropping.
- Experiment #3 is carried out on raw datasets with no noise removal but with image cropping.
- Experiment #4 is carried out on datasets with noise removal and image cropping.

6) *Evaluation*

The results are evaluated using 8-fold cross validation. Then, the performance is expressed in terms of Accuracy, SE and SP, that is calculated using the following formulae:

$$SP = \frac{TN}{FP+TN} \tag{1}$$

$where,$
$T_N = True\ Negative$
$F_P = False\ Positive$

$$SE = \frac{T_p}{T_p + F_N} \tag{2}$$

$where,$
$T_P = True\ Positive$
$F_N = False\ Negative$

$$ACC = \frac{(TP+TN)}{(TP+TN+FN+FP)} \tag{3}$$

- SP – The ability of a test to correctly identify those without DME disease.
- SE – The ability of a test to correctly identify those with DME disease.

IV. Results and Discussions

Table 1 shows the classification performance on four experiments.

TABLE 1

Classification performance in terms of accuracy (ACC), specificity (SP) and sensistivity (SE) in (%)[14]

Pre-processing		**Fully Connected Layer**	**ACC**	**SP**	**SE**
Noise removal	**Image Cropping**				
No	No	FC6	93.92	92.22	95.66
		FC7	88.21	92.78	83.64
		FC8	81.23	85.45	77.00
Yes	No	FC6	94.70	92.38	97.02
		FC7	92.87	90.63	86.18
		FC8	87.30	95.12	88.43
No	Yes	FC6	94.41	95.75	93.07
		FC7	83.96	90.28	89.26
		FC8	82.96	76.66	77.64
Yes	Yes	FC6	**98.85**	**98.39**	**99.32**
		FC7	**96.83**	**96.24**	**97.41**
		FC8	**96.07**	**94.48**	**97.66**

For all 4 datasets, FC6 shows the highest accuracy compared to FC7 and FC8. Meanwhile, denoised data shows significant improvement for both cropped and uncropped data. Nevertheless, denoised and cropped data has the highest SP and SE compared to other datasets. However, there are certain inconsistency of SP and SE, whereby, lower FCL have higher SP and SE compared to higher FC layer. This might due to the pre-trained weights of the FC layers.

TABLE 2

Comparison of Classification Performance between other researchesrs in terms of SP and SE in (%) [8]

Ref	**Srinivasan *et al.***	**Alsaih *et al.***	**AlexNet**
Pre-processing	Noise Removal Flattening Cropping	Noise Removal Flattening Cropping	Noise Removal Cropping
SE	68.8	75.0	94.48
SP	93.8	87.5	97.66

Srinivasan *et al.* [15] and Alsaih *et al.* [16] researches are chosen for comparison because of similar pre-processing methods i.e. BM3D filtering and image classifier i.e. SVM [8]. Both Srinivasan *et al.* and Alsaih *et al.* [15, 16] classify three categories from the OCT volumes i.e. DME, Age-related Macular Degenration (AMD) and normal patients. Srinivasan *et al*. [15] uses HOG feature extractor. In fact, Alsaih *et al.* [16] is an extension of Srinivasan *et al*. research and they added Local Binary Patterns (LBP) to HOG to extract more features and PCA as feature representation to reduce the number of dimension. The results were clearly promising as the SE and SP of Alsaih *et al.* [16] is more consistent and relatively higher than the previous work of Srinivasan *et al.*

However, when AlexNet was introduced, at the end of FC8, the SE and SP is more than 90%. The most significant improvement is the increase in SE performance of 25.68% from Srinivasan *et al.*'s [15] research and 19.48% from Alsaih *et al.*'s [16] research. The increase in SP performance, meanwhile are between 1-11%. Nevertheless, using deep learning model and CNN as the feature extractor, the classification performance increases tremendously.

A. Future Works

For future work, feature reduction using Principal Component Analysis (PCA) or Bag of Words (BoW) could be inserted before classification to reduce the number of dimensions. Then, other deep learning models could be explored with deeper layers e.g. GoogleNet and the results could be combined in a voting scheme. Other than that, one could also try using Leave-One-Out Cross-Validation (LOO-CV) strategy as validation to provide more promising results [9] on a volume based approach. Lastly, fine-tuning technique should be explored to readjust the pre-trained weights of the model to improve classification performance of the current results.

V. Conclusion

In conclusion, the development of OCT which provides high resolution of retinal images for DME detection plus the adaptation of deep learning has proven to improve image classification with high performance of more than 90%. Deep learning application on DME detection using AlexNet has increase in SE performance of more than 20% compared to both researches conducted by Srinivasan *et al.* and Alsaih *et al.* This opens to a new, simple and effective method for early DME detection to aid ophthalmologists in biomedical technologies. For future works, one can look into image classification of other deep learning model to compare the performance on models with more convolutional layers, include feature reduction process, use LOPO-CV for data validation and introduce fine-tuning technique to improve the classification performance of the current results as research findings have shown that using CNN as feature extractor does give almost 100% accuracy [14].

References

[1] R. Lee, T. Y. Wong, and C. Sabanayagam, "Epidemiology of diabetic retinopathy, diabetic macular edema and related vision loss," *Eye and Vision,* vol. 2, p. 17, 09/30. 08/11/received. 09/01/accepted 2015.

[2] B. Cense, N. A. Nassif, T. C. Chen, M. C. Pierce, S.-H. Yun, B. H. Park*, et al.*, "Ultrahigh-resolution high-speed retinal imaging using spectral-domain optical coherence tomography," *Optics Express,* vol. 12, pp. 2435-2447, 2004/05/31 2004.

[3] M. Sonka and M. D. Abràmoff, "Quantitative analysis of retinal OCT," *Medical Image Analysis,* vol. 33, pp. 165-169, 2016.

[4] M. Copeland. What's the Difference Between Artificial Intelligence, Machine Learning, and Deep Learning? [Online]. Available: https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/

[5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature,* vol. 521, pp. 436-444, 2015.

[6] Y. Taigman, M. Yang, M. A. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1701-1708.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097-1105.

[8] J. Massich, M. Rastgoo, G. Lemaître, C. Cheung, T. Wong, D. Sidibé*, et al.*, "Classifying DME vs Normal SD-OCT volumes: A review," in *23rd International Conference on Pattern Recognition*, 2016.

[9] G. Lemaître, M. Rastgoo, J. Massich, S. Sankar, F. Mériaudeau, and D. Sidibé, "Classification of sd-oct volumes with lbp: Application to dme detection," 2015.

[10] M. Lebrun, "An Analysis and Implementation of the BM3D Image Denoising Method," 2014.

[11] D. Fu, H. Tong, S. Zheng, L. Luo, F. Gao, and J. Minar, "Retinal status analysis method based on feature extraction and quantitative grading in OCT images," *BioMedical Engineering OnLine,* vol. 15, p. 87, 2016.

[12] S. P. K. Karri, D. Chakraborty, and J. Chatterjee, "Transfer learning based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration," *Biomedical Optics Express,* vol. 8, pp. 579-592, 2017/02/01 2017.

[13] S. J. Chiu, X. T. Li, P. Nicholas, C. A. Toth, J. A. Izatt, and S. Farsiu, "Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation," *Optics Express,* vol. 18, pp. 19413-19428, 2010/08/30 2010.

[14] Image Category Classification Using Deep Learning [Online]. Available: https://www.mathworks.com/help/vision/examples/image-category-classification-using-deep-learning.html

[15] P. P. Srinivasan, L. A. Kim, P. S. Mettu, S. W. Cousins, G. M. Comer, J. A. Izatt*, et al.*, "Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images," *Biomedical Optics Express,* vol. 5, pp. 3568-3577, 2014/10/01 2014.

[16] K. Alsaih, G. Lemaître, J. M. Vall, M. Rastgoo, D. Sidibé, T. Y. Wong*, et al.*, "Classification of SD-OCT volumes with multi pyramids, LBP and HOG descriptors: application to DME detections," in *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, 2016, pp. 1344-1347.